UNIVERSIDAD
Panamericana

THE WAR SIMULATOR

COM 139: SIMULATION & VISUALIZATION [1]

1 INTRODUCTION

War is an old and very complicated endevour. All kinds of people, from politician to philosophers have tried to decipher the reasons why as a species we have always been inclined to exercise war against each others. Their findings, some more well thought than others, have been wide ranging, from the obvious to the outrageous. For instance, a natural cause of a conflict between groups could be resources, the lack of them by one group and the unwillingness of the other to trade/share. While on the more outrageous wing, we find conflicts that have been started because a group's ideas on how their way of worshiping their god(s) is better than the others'.
Although this is a very simplistic way of observing at war, the truth is that war is a very complex situation, where the "motive" is just really an excuse, while the "real motive" seems a bit harder to device. The very act of going to war produces a series of chain reactions that sets a nation completely off their stable state, where sometimes it takes a nation to complete anhilation or the greater glory. It is known that some of the fastest times for technological development are war times, but it is unavoidable to think on all the suffering that comes with it. Throughout our history we have witnessed all kinds of situations, where great nations have crumbled to pieces after a war or humble ones that have jumped to lead the global stage.
On this work we will create a war simulator, that will try to ilustrate how conflict can brew within a nation or within two or more neighboring nations. This is by no means an exhaustive attempt to reproduce real war situations, but we will do our best to make it as close to reality as we can, based on our limited knowledge and understanding. We will divide the work in 3 distinct sections that will inform us on our understanding of the situation. The first part will require an extensive data analysis task where we will learn about the history of human conflicts in the modern era. The second part will require us to create a simulator with preset rules that will simplify the interactions between nations. The simulator will generate data that we will use in the third part, where we will need to visualize the behavior of the multiple simulations in the attempt to observe if we can shed some further light into the reason on why conflicts occurr.
We will simplify this simulator a lot, thus we will not consider:

- International cooperation or nation blocks
- Different types of warfare. We will only consider 1-on-1 neighboring ground attacks, thus things such as aerial or naval warfare is out of the scope; although a potential extension if you want to further work on this simulator after class.
- No weapons of mass destruction
- Although education will be considered in our input readings, we will not consider the evolution of education into technological superiority for the conflict. We will integrate this factor into the capability of war index of a nation.

[1] *Engeneering Faculty, Universidad Panamericana, Guadalajara, México*

- Regions within a country. This means that we will not be able to cripple a nation's hability to fight by destroying their industry or such things.
- Population age or gender. Thus as long as the nation has population, it has able soldiers to fight and produce.
- Social classes. Every individual of the population belongs to the same social group and has the same income as the rest. The information available can also help us model the social interactions, but that will not be considered for us, but another potential extension.
- Education is equal for all.

1.1 General behavior of the simulation

The simulation will have a Game-of-Life-like behavior, where neighboring nations will start conflict or coexist peacefully based on the indicators of each nation; inner conflict will also be achievable. Initial parameters will be described later, but they will setup the board to allow the simulator rules to evolve to potentially generate conflict between/within nations. The game will be restricted by a maximum number of years the simulation will evaluate, as well as the number of nations that will take place in the simulation itself, maximum 7. After the simulation has concluded, we will analyze the behavior of the world under the given conditions was a more peaceful one.

2 DATA ANALYSIS

To learn about human conflict we will analyze the Correlates of War Project database, the Gapminder project database and other additional complementary sources as needed.
These data sources gather a wealth of information regarding the last 200 or so years of wars as well as some crucial information regarding the economics, politics and population stats for all the nations in the world. This information will help us profile the initial parameters for our nations and how we can push a nation to go to war or not. Some of the questions we can ask from the data that could help us guide our analysis could be, but are not limited to:

- Identify the regions of the world where conflict is more common
- Measure the length of wars
- Measure the number of casualties by war.
- Identify the number of conflicts that evolved from local to international and viceversa.
- How many wars each country has been part of?
- How many wars each country has started?
- What are the nations that have been attacked the most?
- The average peace time between wars
- Can I do a timeline of wars?
- What is the average income of the nations involved in war before and after?
- What is the life expectancy of nations while there is no war and before, during and after a war?
- What is the literacy rate of the population? Is this a factor in the aggresiveness of a nation?

- What is the population of the nations in conflict, with no war, before, during and after war?
- What is the military spending by region? by % of GDP?
- ...

All these questions will guide you to identify the distribution of the values that will serve to initialize the simulation.

2.1 Sources

The following links lead to some sources you may use to properly simulate your input data.

- https://www.kaggle.com/datasets/sdorius/globses
- https://correlatesofwar.org/data-sets/COW-war
- https://www.gapminder.org/data/
- https://en.wikipedia.org/wiki/List_of_countries_by_military_expenditures

2.2 Data preparation

For simplifying purposes we will group our data by world regions. The 7 regions we will use are: Europe, Africa, Asia, N.America, S.America, Oceania & S.E. Asia and Middle East.
All the data from countries in each region should be grouped as such. Once we have gathered all the data we will identify the distribution for the following indicators:

- Territory
- Population
- Population growth rate
- Income
- Literacy rate
- Military spending by % of GDP

We need to identify what type of distribution each region resembles. We will need to identify the mean and the standard deviation, since these two measurements will help us device the threashold values that will trigger events as our simulation evolves.

3 THE SIMULATOR

This part of the project will be based on the previous one. You are free to use any framework that you see fit, or you can use SimPy as we used in class. This piece of software should consider the findings in the previous section to select properly the distributions of the input data as well as the required parameters to really represent the data observed in the data exploration. You should be able to run it quickly and produce detailed output so you can later create the most pertinent visualizations to tell your story.
The rules described below are a proposal of the workings of the simulator, this is a work in progress, so if at any point you feel that something is not right, feel free to adjusted as you see fit.

The world will be represented by an nxn grid and every cell will represent a square unit of territory. It is important to have a clear count of the territory since that will be base for the calculation of the population and total wealth. This will also be modified as a nation goes into war with others.

3.1 Initialize

From the distributions and values obteined previously, you will need to initialize every country with a random variable extracted from the distribution for each of the following properties:

1. Territory
2. Population
3. Population growth
4. Income per capita
5. Literacy Rate
6. Military spending by % of GDP

From a random normaly distributed variable you will get:

1. Goverment Rate: A value between 0 - 1 that will depict how strong a goverment is; 0 is weakness, while 1 is total confidence and strength. The stronger it is, the more represive it will get, the weaker it is, the harder it will be to deal with internal issues.

Once you have these values, you can calculate:

- Total Income: Population * Income per capita

3.2 Growth times

For every month of the simulation that runs you will update the population by a 1/12 of the population growth rate and then update the country's total income.

3.3 War mechanics

Every day simulated neighboring squares will be evaluated to check if any trigger has been activated. The squares are first checked sideways from left to right, and then from top to bottom, no diagonal check. A square that has already battled against a lateral enemy territory should not battle again against a top-bottom facing territory. Every time there is a conflict a war will be declared in the ledger and all its actions should be registered for later evaluation.

3.4 Triggers

For this simulation we will just focus on 2 parameters to become triggers: income and population density. To reach trigger situations we will use world distributions, that is, we will use the income and population density of the whole world in order to build the distribution and get the mean and the standard deviation. It is suggested to have a system configuration named INCOME_THREASHOLD and POP_THREASHOLD to fine tune the triggering mechanics.

3.4.1 Income trigger

When 2 neighboring countries reach a difference of INCOME_THREASHOLD times the Std. Deviation of the world income a conflict will arise WITHIN a country. But if the country has a strong goverment the conflict can be withheld. But this is a clear sign that in the next elections probably conflcit will result in a civil war. Nevertheless, if there is a neighboring country whose **total income** is within one Std. Deviation of own's, then a Foreign country will be launched.

This behavior is modeled to simulate social unrest due to economic hardship, and how some authoritarian goverments can hinder that desire for change, as well as how some take evasive action and defer the attention from their incompetence and prefer to go to conflict with a third party instead of solving the issue at hand.

3.4.2 Population trigger

When the population density reaches one Std. Deviation above the median of the region's a conflict will be triggered in order to look for further territory. The country that will start the war will select it's neighbor with the lowest lower income to attack. If the country at conflict happens to be the poorest of the region and no other country has less income than itself, then Internal conflict will arise.

3.5 Conflict

The way that conflict will be addressed is similar to the way the Game of Life works. Every square area in contact with a square of the battling country will have to undergo a battle. The way that a battle is decided is by the use of the Capability of War Index (CWI). In any case, the block that has the bigger value will beat the other one. Refer to the Capability of War Index section for further detail on how it is calculated and used.

3.5.1 Capability of War Index

On every square combat we will calculate the Capability of War Index (CWI). The CWI is a made up value that will combine the information we have and will be used to decide the outcome of a battle. The formula is:

$$CWI = (TI)(GR)(LR)(MS)(LK)$$

- TI - Country's Total Income
- GR - Goverment Rate
- LR - Literacy Rate
- MS - Military Spending
- LK - Luck!*

* **Luck** is a system setting, that will represent a boost for the group fighting. It is suggested that it should be a value below 0.1 in a normaly distributed variable to activate the boost. And the boost is suggested to be 1.3, although this should also a system variable, thus its prone to be played with to see what gets the best results. A higher value will make the "lucky" side stronger against the rival.

3.5.2 Internal conflict

When a country is fighting a foreign country the mechanics are pretty obvious, one country vs the other one. But when the conflict is within it is not that simple. When an internal conflict is brewing, this are the rules:

- A random walk of the country's territory will start, where half of the squares representing the country should be visited.
- On every square visited a Random Normal variable should be evaluated. If the value is 0.5 or more, then a battle should take place. Else, no further action will be taken.
- If a battle shoudl start, we will calculate the CWI, but we will modify the Luck factor to be double of the suggested one for the insurgency group.

- If at the end of this random walk there is no victorious block for the insurgents, then the civil war is deterred and everything remains the same for another cycle, except casualties will be taken.
- * If on the other hand, there are some territories where insurgents succeeded, then the mechanics will repeat with the neighboring territories of these.
- This process will repeat until there are no more insurgent territories or the insurgents have achieved one third of the country's territory if the goverment is < 0.4 or two thirds of the territory if the goverment is > 0.4.
- If the status quo succeeds, then everything is back to normal, else a new goverment will arise via the Goverment Change procedure described later, but this goverment will have a 0.2 extra value, to simulate the population support.
- The conflict will stop completely if casualties reach 0.4 percent of the population that existed when the conflict started. In such case a new Goverment will be elected and the country will stay as is.

* Insurgents will keep the bonus **Luck** factor to simulate the social support.

3.5.3 Casualties

Once the result of a battle has been decided we have to calculate the death toll on both sides. This measurement is completely made up, so it is also prone to be fixed if we see this is not working properly. The proposal to calculate the Death Factor (DF) is the following:

$$DF = (CWI-)/(CWI+)$$

- CWI- : CWI of loosing side
- CWI+ : CWI of winner side

* DF should be a value between 0 and 1.
Once we have calculated the DF, we will multiply it by the Population Density (PD) of the **loosing** country and substract them from the country's population:

$$Casualties- = (DF)(PD-)$$

The winner side should also have casualties, though they should be less. For this we will use the following formula:

$$Casualties+ = (1 - DF)(DF)(PD+)$$

When the conflict is internal you should use twice the value of the winning side (Casualties+).
The lost of population due to war will be reflected in the country's income, since income is calculated by individual.

3.5.4 Outcomes

When a side looses the territory that was in dispute becomes part of the winning side, thus updates in territory have to be made. The population that is not considered casualties is assumed to have fled deeper inside its country and no captured population is ever considered.
If there ever is a draw on the hostilities, then half of the population of the disputed area on both sides will be deducted from their populations. No territory is lost by neither.

3.5.5 *End of conflict*

The rules for inner state conflict have already been stated above.
For inter state conflict, the event will only finish when:

- A state has conquered 50% of the enemy's territory, at which point the rest of the country's territory is annexed by the winner side, taking in all the surviving population as well. No further merge is done.
- The income of one of the countries is 3 Std. Deviations from its adversary. At which point there is a cease fire in all territories, the winning side keeps the conquered territory and the loosing side dwells in shame and calls for new elections.
- There is a stale state for 15 days. The stale state means that there has not been a net gain or loss above the 1% of the territory on either side. At this point, there is a cease fire and new elections are called on both countries.

3.6 Peace times

It is suggested that another system variable is set to reevaluate the parameters of a country periodically. It is suggested to do so every 10 years. For this simulation, if a country has not been in war within that period will achieve the following:

- + 1/10 of the Std. Deviation (σ) of the literacy rate of the country
- + 1/8 of the Std. Deviation (σ) of the income of the country
- if the literacy rate is:
  - LR < 0.4 : + 1/3 of population growth
  - 0.4 < LR < 0.7 : + 1/6 population growth
  - 0.7 < LR < 0.9 : - 1/6 population growth
  - 0.9 < LR : - 1/3 of population growth

These changes are in place to reflect social progress and prosperity in a nation that has stayed out of conflict.

3.7 Goverment changes

Another thing we can do to simulate changes is the change of goverment. This can be done either by country or by a system variable, but I suggest timeframes of 5 years. Let's remember that the Goverment variable is just an indicator of how strong/weak a goverment is. This factor affects the outcome of a battle, but also the tendency of start conflict within the country. As stated in the initialization, a normally distributed variable 0-1 should be used, where > 0.7 means a strong country and < 0.3 means a weark goverment that is prone to fall quickly.

4 THE DATA VISUALIZER

Data visualization is simply the visual representation of data. This might be basic charts and tables that are generated from a spreadsheet or it could go well beyond those modalities to include any use of shapes, color, and sizing to draw visual focus to data findings. Bottom line, data visualization is about communicating the substance of your metrics in a visual way. Data visualization can certainly be used to tell a story

at the slide level. It can provide context, elevate and draw attention to key insights and lead to action.
Graphic displays are often very effective at communicating information, but the opposite is also true. Two important reasons for this state of affairs are that graphics can be produced with a few clicks of the mouse without any thought and the design of graphics is not taken seriously in many scientific textbooks. Some people seem to think that preparing good graphics is just a matter of common sense, while others believe that preparing graphics is a low-level task, not appropriate for scientific attention.
However, the real magic happens when data visualization is driven by storytelling at the slide level and story level. Contrary to popular belief, data storytelling is not simply data visualization, analytics reporting, or a handful of stats sitting in a PowerPoint somewhere. Data storytelling is the general term we use to describe the full act of gathering data, extracting insights, and translating those insights into a story. Data storytelling is the blending of two worlds: hard data and human communication. It's a compelling narrative crafted around and anchored by compelling data, used to guide decision-making, reveal an interesting trend, or provide valuable information to your audience.
Storytelling with data differs from data visualization because it requires communicators to offer a larger, holistic, view of their message. You must focus first on your audience and structure a larger message before any visuals are rendered. You must identify from the start:

- What do I want my audience to know or do with the data I am presenting?
- How will I structure a narrative that leads to desired action?
- How is my data helping drive a decision?

There is no understating how important it is for all presented data to have a purpose. Every piece of data you include should further this purpose – or it should be left out.
Good data storytelling isn't just slapping together a few charts. It requires several ingredients.

- Good data: The raw data your company collects.
- Synthesis: Identifying the story the data is telling.
- Narrative: Translating those data insights into a clear story.
- Data visualization: Communicating data insights through visualizations that make it easy to "see" the story you're telling (e.g., the increase in seasonal sales or the percentage of satisfied customers).

To do data storytelling well, you need every piece of the puzzle.
We all know that telling a story with data gives it meaning. But many of us wonder: How do we actually do this? Well, the work begins before you sit down at your computer to build that chart or table. A good data storyteller always begins with their audience and establishes their key message:

- What am I trying to achieve with the data I choose to display?
- Who is my audience? What do they care about? What level of data detail will they likely expect or appreciate?
- What is my BIG Idea – the one thing I want my audience to know or do with my data?

Storytellers also determine: "Is the data I am choosing moving forward the story I want to tell?" If the answer is yes, it is probably safe to include it. At that point, you are ready to open your computer and begin to build your visuals.

5 WHAT MAKES A GREAT DATA STORY?

How can you tell a data story that that helps people make better decisions? Let's consider what makes a data story effective.

5.1 It needs to be relevant

This means the content (including copy and visuals) needs to fit with the audience's current level of knowledge and it needs to help them reach a goal of some kind. Maybe your audience is internal, like a presentation to leadership about the need to invest in a specific strategy or tactic. Or maybe they are external, such as a campaign to persuade customers to try your solution. Either way, think about what matters to them. The best stories speak to people, and the more specific the person, the better.

5.2 It needs to include good data

This means the data should be from a reputable source and/or collected in a way that truly represents what's needed to tell a truthful story. Publicly available data from government entities, intergovernmental organizations, academic researchers, and established leaders in analysis are not only more accessible, they are also transparent and verifiable.
But having a good data source alone isn't the only thing to consider. For example, you wouldn't use data about positive tests for COVID-19 in New York as your only source of data to talk about how COVID-19 is impacting the world, right? Since data is the center point, these stories are not about what you think your audience should hear, they are about sharing what the objective data says. This is markedly different from other stories we might tell. The data used should help you tell the honest truth. It should align with the audience's needs and support them in learning exactly what they need to know to make a key decision.

5.3 There needs to be a clear narrative

When it comes to storytelling, we are all used to a traditional story arc with a beginning, middle, and end. For data stories, this usually means you need an introduction to the topic before you dive into the data. You also need to conclude with a specific call to action—this is another thing that makes a data story distinctly different from a straightforward report. Also, if your audience is not experts, it's important to use plain language so you don't lose them in wonky jargon or confusing acronyms.

5.4 It should include intentional visuals

What does intentional mean? It means whether you use photos, graphs, and/or charts, the visuals you use should help your audience easily understand what the data means.
Above all, the visuals you include should be:

1. Appropriate for the data
2. Well-labeled
3. Legible
4. Not misleading

Great data stories pay attention to details like use of color and imagery, including considerations related to accessibility and diversity.

6 STEPS TO CREATING A GREAT DATA STORY

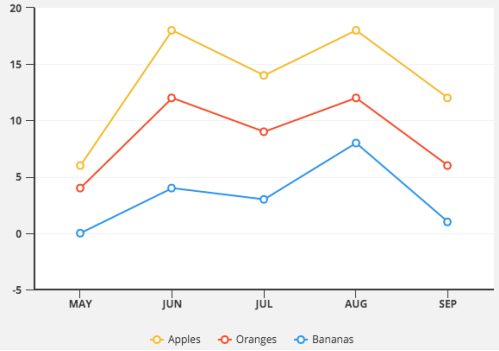
Lets now walk through a simple example so you can see for yourself how great data stories are made, from scratch.

6.1 Start by listening to your audience

First, you need to identify who they are, and then you can talk to them and perhaps do additional research to find out what they most care about, what their goals are, what they currently know, what decisions need to be made and what additional knowledge might help them make the decisions that will help them reach their goals.
Here's an example of how we can create a data story:

> My audience is people working in online education who want to understand how the COVID-19 pandemic impacted people's desire to learn more about virtual learning opportunities.

We will create a data story that will help them make decisions about whether they should continue to develop these opportunities. We can use data from Google Trends for 2020 in the U.S.



6.2 Pinpoint the data that matters

Knowing your audience will help you know what data to look for and include. You might use quantitative data, such as about revenue, change over time, or number of people impacted, or you might use qualitative data, such as processes, systems, or quotes.
For example, this simple line chart template is easy to edit to show change over time.
Example: We review data for Google's Year in Search 2020 data, and we see that they have outlined the top 10 searches that included "virtual." Out of the 10 listed, three are related to online education.
We could look for additional related data, for example that the Department of Education, school districts, or other reputable data sources collected, but since we are going to keep this example simple, we'll only include this data here

6.3 Outline the story arc

Once we have our data, we can explore some possibilities. Do we want to show that three of the top 10 were education-related, or do we want to dive deep into the three that matter most, or both? We explore the data, usually looking for where there might be three parts to a story.
Since this data is annual, there is a natural story arc, especially since the school year has a predictable rhythm. The beginning of the story could be about how the year started in a traditional way, how COVID-19 disrupted studies in the Spring, and about how Fall's back-to-school looked completely different.

6.4 Create a draft design

With our story arc in hand, we can think about what sorts of design layouts or compositions might work best. We want to get a better idea of what will work visually, so we often sketch out, by hand, some different layouts and compositions. For instance, we can look through numerous data visualization templates to get some ideas of what might work. We know we want to look for some that will help us show change over time, we could just highlight three main ideas or just three important numbers.
Once we have selected a layout or template, we can modify and add visuals that show the story as well as copy that helps tell the story. We can add text boxes to the infographic for a title and graph labels to help the audience understand the graphs.

6.5 Assess your blind spots

Now that we have a draft, we can pause and reflect on any assumptions we might be making about the audience, about the data, about the story we are telling, and/or about the visuals we've created. It's super easy to skip this step, but if you want to create a great data story, it's worth taking a bit of time for reflection. The very best way to do this reflection is by sharing your draft with a colleague, or even better a person who identifies with your audience.
For example, you can ask your team or a colleague to offer their thoughts on the draft shown and allow them to give me their comments. This can help us recognize that we need to make it clear to the audience that the numbers are much higher/lower than they were expected, and that we can use a title to suggest there is a trend that the audience should address.

6.6 Polish up the visuals and share

Now we can make some changes to make sure my story makes the most sense to the audience, and we can add some details if we want, so long as they don't take away from the story.

The best stories stimulate dialogue, so you can expect to get more feedback when you start sharing. You can use this feedback to improve your skills and perhaps to think of other data stories you might create in the future.

7 THE TASK

As a team, you will have to develop the 3 parts of the project: Data analysis, Simulator and Data visualizer. Thus, each part will have different requirements:

7.1 Data analysis

1. A piece of code that automatically process new input data and displays the required information.
2. This should include graphics displaying the work done in order to reach the distributions.
3. This work should be either printed or in PDF with an extension long enough to guarantee that all the requested distributions were analyzed in detail.
4. This part of the team should be ready to answer some basic questions on how the analysis was done and how some data is obtained.

7.2 Simulator

- The simulator running in full with easy to configure files to modify the behavior of the simulation
- Should produce output files that will later become the input for the visualization part.

7.3 Data Visualizer

- Should have multiple graphs that display the results obtained by the simulator, it should allow the data to be sorted, filtered and grouped to change the way that visualizations are displayed.
- It should be running in a live web server with public access.

- The list of components to be displayed can be defined by the team, but they should not be less than 5 fully functional graphs with filter and grouping capabilities.

Each of the points requested above have a 20% weight for the grade of this project. The optional requirement will add a 10% to your final grade.

7.4 What to deliver

- Source code of the simulator ready to execute with clear instructions on how to install, configure and run.
- Printed work of the analysis
- Source code of the visualization tool developed to do the story telling of the project findings.
- Development log. This is just a document with the explanation of the challenges you faced and how you solved them. It is perfectly OK to state that some of the challenges were not solved. If you refer to other sources to solve a problem, please don't forget to cite them properly.
- At least 5 stories that showcase and explain what information is getting displayed.
- Presentation used for the final project presentation.